LEGIBILITY NOTICE

A major purpose of the Technical Information Center is to provide the broadest dissemination possible of information contained in DOE's Research and Development Reports to business, industry, the academic community, and federal, state and local governments.

Although a small portion of this report is not reproducible, it is being made available to expedite the availability of information on the research discussed herein.

LA-UR 90-515

February 1990

Los Alamos National Laboratory is operated by the University of California for the United States Department of Energy under contract W-7405-ENG-36

LA-UR--90-515

DE90 007539

TITLE: "Application of Neural Nets to System Identification and Bifurcation Analysis of Real World Experimental Data"

AUTHOR(S): R.A. Admoaitis, Princeton University
R.M. Farber, Los Alamos National Laboratory
J.L. Hudson , University of Virginia
I.G. Kevrekidis, Princeton University
M. Kube, University of Virginia
A.S. Lapedes, Los Alamos National Laboratory

SUBMITTED TO: International Conference on Neural Networks Proceedings
Lyons, France

Los Alamos Los Alamos National Laboratory
Los Alamos, New Mexico 87545

FORM NO 836 R4
ST NO 2629 5/81

Application of Neural Nets to System Identification and Bifurcation Analysis of Real World Experimental Data

R.A. Adomaitis†, R.M. Farber‡, J.L. Hudson*, I.G. Kevrekidis†, M. Kube*, and A.S. Lapedes‡

**Department of Chemical Engineering*
University of Virginia
Charlottesville, VA 22903-2442
† Department of Chemical Engineering
Princeton University
Princeton, NJ 08544
‡ Theoretical Division
Los Alamos National Laboratory
Los Alamos, NM 87544

Abstract

We report results on the use of neural nets, and the closely related "radial basis nets", to analyze experimental time series from electro- chemical systems. We show how the nets may be used to derive a map that describes the nonlinear system, and how reserving an extra "input line" of the network allows one to learn the system behavior dependent on a control variable. "Pruning" of the network after training appears to result in elimination of spurious connection weights and enhanced predictive accuracy.

Subsequent analysis of the learned map using techniques of bifurcation theory allows both nonlinear system identification and accurate and efficient predictions of long-term system behavior. The electrochemical system that was used involved the electrodissolution of copper in phosphoric acid. This system exhibits interesting low dimensional dynamics such transitions from steady state to oscillatory behavior and from period-one to period-two oscillations. This analysis provides an example of methodology that can be fruitful in understanding systems for which no adequate phenomenological model exists, or for which predictions of system behavior given a large scale, complicated model is inherently impractical.

1 Introduction

Neural networks have recently been shown to be useful for nonlinear signal processing tasks such as the prediction of nonlinear time series [11] [12] . They have superior predictive accuracy when compared to the conventional approach to this problem involving Volterra-Wiener expansions, and are competitive in accuracy with other new (and related) approaches (e.g. [7] [8]). There is also an intense effort to build neural nets in hardware in order to take advantage of the speed inherent in massively parallel VLSI analog hardware. Given the new abilites of neural networks, and the speed possible in hardware implementations, it becomes very interesting to ask what range of useful tasks such networks could perform in the realm of nonlinear signal processing. In this paper we discuss some recent results we have obtained in which a neural net is "trained" to model the behavior of a real system dependent on a control parameter. After training, the net can be used not only to predict the future history of a time series from the system, but can also accurately predict qualitative changes in system behavior as the control parameter is changed. Furthermore, bifurcation analysis can be applied to the neural net after it is trained to perform a type of nonlinear system identification.

In the past few years there has been considerable interest in the long range prediction of chaotic time series using neural and other methods. Most of the this literature deals with simulated data, produced for example, by numerically integrating a chaotic nonlinear differential equation. In this paper we utilize neural networks and radial basis networks to analyze real world experimental (electrochemical) data. Furthermore, we wish to do more than predict the future values of some time series from the system. Instead we wish to be able to predict qualitative behavior changes as a control parameter is varied and to attempt nonlinear system identification. In this early study, we limit the analysis to transitions from steady state to periodic oscillations and from single-loop to double-loop oscillations. In subsequent papers we also plan to analyze transitions to chaos. Our goal is to investigate the hypothesis that long-term prediction of chaotic time series, while intrinsically interesting, may not be the most practical application of these new methods. Instead, the ability to *qualitatively* predict what occurs in control parameter space, especially changes of stability and bifurcations, is exceedingly relevant for important problems such as adaptive control of nonlinear systems.

2 Electrochemical Experiments

For our pilot study we chose data from chemical reactions. Chemical reactors often operate under unsteady conditions. This unsteady behavior can result from variations in an input, such as a feed concentration, or can be caused by kinetic or thermal instabilities in the reactor system itself. In dealing with experiments under such conditions, one obtains a transient signal, e.g., a measurement of concentration, potential, or temperature as a function of time. One way of interpreting such signals

is by comparison to fundamental models, obtained using established theories of the process. These models can then be used in characterization of the process, selection of optimal operating conditions, control etc. However, one is often faced with processes for which fundamental models are tentative, nonexistent, or extremely difficult to analyze because of their size and complexity. On the other hand, the experimentally observed dynamic behavior may often be low-dimensional, suggesting that a small set of ordinary differential equations could provide an accurate model of the system, at least over the operating regime of interest.

Electrochemical reactions are known to exhibit interesting dynamic behavior. For example, the electrodissolution of copper in acidic chloride solutions is known to undergo several well-defined low-dimensional transitions, such as breaking of tori, transitions to homoclinic behavior, and period-doubling of tori (e.g. [3] [2] [1]) . The dynamics of these systems have been investigated using several methods from nonlinear dynamics, such as Poincare' maps, dimension calculations, and attractor reconstruction through time-delays (the latter proposed by Packard et. al. [13] and by Takens [15]). We apply neural network methods to time-series obtained from this latter system. The experiments are performed under potentiostatic conditions, and the current is measured as a function of time. All data were collected at 1000 H_z. The data consisted of times series of the current (corresponding to the net rate of dissolution) for a range of parameter values (the operating parameter is the applied potential).

3 Nonlinear Signal Processing

3.1 Neural Nets

The use of neural nets for prediction of time series has already been described elsewhere citeLapedes87a [12] . Therefore in this section we briefly review the methodology and also describe the relation of radial basis functions to neural nets.

We wish to take a set of values of x() (where x(t) is an experimentally measured times series) at discrete times in some time window containing times less than t, and use these values to accurately predict x(t + P), where P is some prediction time step into the future. That is, we wish to construct a map from the past values of x(t) to a future value of x(). One may also include in the domain of the map a control parameter, and hence determine the effect of the control parameter on the range of future behavior.

Our goal is to use backpropagation, and a neural net, to construct a function

$$0(t+P) = f\,(I_1(t), I_2(t-\Delta) \ldots I_m(t-m\Delta), E) \qquad (1)$$

where 0(t + P) is the output of a single neuron in the Output layer, and $I_1 \rightarrow I_m$ are input neurons that take on values $x(t), x(t-\Delta) \ldots x(t-m\Delta)$, where Δ is a time delay. E is the value of the control parameter and has a separate input line. O(t + P) takes on the value x(t + P). We chose the standard feedforward network configuration

with synaptic weights T_{ij}, thresholds Θ_i, and two hidden layers inserted between the input and output layers. The hidden layers use the usual sigmoidal transfer function

$$O_i = g(\sum_j T_{ij} I_j + \theta_i) \tag{2}$$

while the output neuron's transfer function is linear.

A training set may be constructed by selecting a set of input values:

$$I_1 = x(t_p) \tag{3}$$

$$I_2 = x(t_p - \Delta) \tag{4}$$

$$I_m = x(t_p - m\Delta) \tag{5}$$

with associated output values 0 = x(t_p + P), for a collection of discrete times that are labelled by t_p. Thus we have a collection of sets of $\{I_1^{(p)}, I_2^{(p)}, \ldots, I_m^{(p)}; 0^{(p)}\}$ to use in training the neural net. This procedure of using delayed sampled values of x(t) can be implemented by using tapped delay lines, just as is normally done in linear signal processing applications, ([16]). The training procedure (backpropagation) is a straightforward nonlinear extension of the linear Widrow Hoff algorithm. The network is trained using nonlinear least squares. This gives values of T_{ij} and θ_i that minimize the sum of squares of the differences between the predicted and measured values of the output sampled over the entire times in the training set. A conjugate gradient method was used for the minimization. The calculations were carried out on the 64,000 processor Connection Machine at Los Alamos. At every iteration of the conjugate gradient method, we simulateously evaluate the prediction of the current net for every point of the training time series. Individual processors evaluate the predicted output for individual points in the time series, thus exploiting the massively parallel architecture of the CM-2. In earlier work ([11] [12]) we showed that the extension of the Widrow Hoff procedure to the Volterra Wiener polynomial expansion in an attempt to perform nonlinear prediction is significantly inferior to the method using sigmoids described above.

Neither m nor Δ has been specifed so far, nor have we given any indication why a formula like Eqn. (1) should work at all. Important work of [13] and [15] shows that for flows evolving to compact attracting manifolds of dimension d_A, that a functional relation like Eqn. (1) does exist, and that m lies in the range $d_A < m+1 < 2d_A + 1$. We choose m=3. Takens provides no information on Δ and we chose Δ from the first minimum of the mutual information [9]. Takens theorem gives no information on the form of f() in Eqn. (1). In earlier work however, we showed that neural nets provide a robust approximating procedure for continuous f().

After the iterative optimization procedure has converged (not necessarily to a global minimum), we have an explicit nonlinear map

$$x_{n+1} = f(x_n, x_{n-1}, \ldots x_{n-m}) \tag{6}$$

where m is the number of delays chosen and x_n is the value of the time series at the n^{th} discrete time sample. By "bootstrapping" the net into the future, this map can be iterated to give

$$x_{n+2} = f(f(x_n, x_{n-1}, \ldots x_{n-m}), x_n, x_{n-1}, \ldots, x_{n-m+1}) \tag{7}$$

or, using an $m+1$ dimensional vector $X_n \equiv (x_n, x_{n-1}, \ldots x_{n-m})^T$ we can define the map

$$X_{n+1} = F(X_n) \tag{8}$$

In this formalism, a measurement or a prediction in the time series becomes a phase point in an $m+1$-dimensional space, with coordinates the present value and the $n-m+1$ previous values. Trajectories in phase space approach an *attractor* as initial transients die out. Comparing the attractors reconstructed using time delays from the original time series, and the attractors resulting from iteration of Eqn 8 is a useful evaluation of the long-term predictive capabilities of the network.

We have incorporated the dependence of the map 8 on the operating parameter by simply reserving one of the neural network inputs for the value of the fixed potential, i.e., we let

$$x_{n+1} = F(x_n, E) \tag{9}$$

Eqn 8 can now be used to perform stability and bifurcation analyses of the system as the operating parameter varies. A comparison of experimental and predicted bifurcation diagrams can give insight into specific dynamic instabilities of the system, and can be useful in designing experiments to elucidate their nature.

3.2 Radial Basis Functions

The purpose of the above analysis is to obtain a nonlinear input-output map which, given the operating parameter values, the present measurement, and some recent history, will predict the state of the system at the next time step. The resulting map can be used for short-term prediction of the time series. Even in the long-term, however, the map can be used to generate the system attractor(s), and perform stability and bifurcation analysis of the system behavior. Sigmoidal neural nets are only one way to obtain this map. In the last few years several techniques for obtaining such maps for nonlinear dynamical systems have been proposed (e.g. [4], [5], [6] [11], [7]). Research efforts have focused mainly on time series prediction, rather than system identification.

A second technique we used for obtaining the map was radial basis function approximation([14]), first used for prediction of time series by Casdagli ([5]). This is an interpolation method that is linear in the relevant weighting coefficients (the analog of the synaptic weights of the neural net), but uses basis functions that are nonlinear in the training data. It may be briefly summarized by saying that it replaces the polynomial basis of Gabor ([10]) with the radial basis of Powell [14]. Radial basis functions are basically localized bumps (in distinction to polynomials that are delocalized and blow up at infinity), and an attempt to approximate a multidimensional

function by radial basis functions may intuitively be viewed as analogous to approximating a one dimensional curve by adding together a number of suitably placed and scaled Gaussian bumps. The localization is presumably why the radial method performs better under iteration than the polynomial basis. The method specifies placing the bump over each training point, which leaves only the amplitude of the bump (which appears linearly) to be determined. Therefore this method does not suffer from a local minima problem although the set of linear equations to be solved may become ill-conditioned for large data sets, resulting in another set of problems for radial basis nets. The radial basis expansion may be written

$$O_i = \sum_j a_j f_j(X_i) \tag{10}$$

where a_j is the linearly appearing adjustable coefficient and $f_j(X_i)$ is the localized radial basis function

$$f_j(X_i) = exp - (|X_i - X_j|^2/\sigma^2) \tag{11}$$

X_j is an $m+1$ dimensional vector formed from the delayed values of the time series. *sigma* controls the falloff of the Gaussian.

The radial basis method is also related to the operation of a two hidden layer backpropagation neural net. As shown in [11] [12] synaptic weights can be found for a two hidden layer back propagation network so that localized bumps can be constructed by the net, and this was suggested as one possible explanation as to how neural nets may perform function approximation. The suggestion was also made that a linear sum of Gaussian bumps might be an effective formalism for prediction. The connection to radial basis functions was unknown to us at that time. Therefore, neural nets can be viewed in one sense as an extension of radial basis nets in which not only the amplitude, but also the width and position, of the bumps may be adjusted. Of course this extra generality pays a price in the nonlinear optimization that is necessary.

There is another type of radial basis function, which although finite at infinity is actually *increasing* locally [14]. This *radially increasing* function may be taken to be simply the distance of a point to a training set point. To maintain localization and avoid blowups at infinity one places the additional constraint that the sum of the coefficients must be zero. This other type of radial basis expansion may be written

$$O_i = a_0 + \sum_j a_j f_j(X_i) \tag{12}$$

where a_j is the linearly appearing adjustable coefficient, a_0 is an additional coefficient specifying the value at infinity, $f_j(X_i)$ is the *increasing* radial basis function

$$f_j(X_i) = |X_i - X_j| \tag{13}$$

and there is the further global constraint that imposes a finite solution at infinity

$$\sum_j a_j = 0. \tag{14}$$

We found (although the point was not intensively investigated) that radially increasing basis functions led to better prediction.

It is also possible to view radial basis function approximation as a kind of network that employs neurons with an non-sigmoidal transfer function. This interpretation should not be confused with the connection to backpropagation mentioned earlier. In this new interpretation, the network is of the simple "perceptron" type and does not involve hidden layers. The output neuron is linear. The required information for our problem is again available on the input layer as the present value, two time-delayed values and the parameter value. However, this information is preprocessed in the following manner: we calculate the distance (in the four dimensional (phase)×(parameter) space) of the present point from a data point. Using k data points from the original time series, we thus calculate k distances. These k distances are the inputs to the k neurons of the single layer of the network. The output (the predicted value of the current at the next time step) is then given by a linear combination of these k distances. Because of the simple, linear dependence of the output on the synaptic weights, training this network is a linear least squares problem, which can be solved using singular value decomposition. Alternatively, one can also use a conjugate gradient method to "train" the network (i.e. determine the coefficients).

3.3 Pruning

Networks with an excess of adjustable parameters may be expected to tune to experimental noise, and therefore although the training set is captured essentially perfectly, the prediction on new data may be poor. This is analogous to the classic "overfit" problem in curve fitting with high degree polynomials. The problem can be particularly severe in radial basis nets, where there are the same number of adjustable parameters as training points. A simple, yet effective, heuristic to eliminate or "prune" unnecessary parameters may be invoked. At a desired point in the training process (we usually chose the end of training) one tests each weight in turn to determine its contribution to overall error. The test is performed by temporarily eliminating that weight (setting the value to 0.0) and calculating the change in the average error over the training set. The weight that causes the least change in the average error is selected as the least important weight. This weight is now permanently set to 0.0, and the network retrained from this configuration. At the end of training the process is again repeated on the new system that has $N-1$ weights. The process terminates when weights are deleted giving unacceptable increases in the average error. Typical reductions in the number of weights is by a factor of 10 (e.g. 5001 weights to 42 weights). In certain runs we found that the "pruned" network seemed to be a better predictor than the unpruned network. This is presumably due to the pruning acting as type of smoothing or noise filter. This effect is under investigation and detailed results will be presented elsewhere.

4 Results

We first describe an example of using the two-hidden layer network approach. This network was trained on two experimental time series, the first at a potential of $788mV$ yielding a single peak oscillation, and the second at a potential of $769mV$ yielding a double peak oscillation. The resulting pruned network was capable of predicting both the single and the double-loop oscillations at the corresponding parameter values. Figures 1a and 1b show the experimental transients at 769 and $788mV$ respectively, including a relatively long startup period. Figures 1c and 1d show two-dimensional projections of the phase portraits reconstructed using two time delays (of 10 and 20 sampling periods respectively). Figures 1e and 1f show the same projection of the phase portraits of the attractors predicted by the pruned neural network. Although not shown in the figure, short-term prediction of the time series is excellent. What we see in this figure is that the network can also capture semi-quantitatively the asymptotic, long-time system behavior (i.e. its attractors).

The behaviors shown in Figures 1a and 1b are far removed in parameter space from each other (i.e. there is no simple bifurcation between the two states). We now discuss a series of transitions which occur in a well-defined manner as the parameter is changed. Figures 2a, b and c show experimentally determined attractors for potential values of $748mV$ (steady state), $758mV$ (a triangular shaped oscillation) and $759mV$ (a double-loop oscillation) respectively. Experimental time series at intermediate parameter values (a total of eight time series between $748mV$ and $763mV$) indicate that the transitions we observe are most probably a Hopf bifurcation to a limit cycle followed by a simple period-doubling. For this set of eight time series, we obtained our best results using the pruned radial-basis network. Again, it is important to note that the parameter is part of the fitting, since the radial distances which constitute the input to the network are calculated in the four-dimensional series in the period-doubled regime ($V = 759mV$). We see again that short term prediction of the time series itself is very good, but prediction accuracy deteriorates as the map gets iterated farther into the future (notice the shift in peaks towards the end of the time series segments shown). Nevertheless, the nature of the two-peaked oscillation (double-looped attractor) is retained here, as it was in Figure 1f. Finally, in Figures 2f and 2g we include the experimental and computational bifurcation diagrams respectively. For the oscillatory branches we plot the maximum and minimum of the oscillations. Since the fitted map is available in closed form, using numerical techniques we can find both stable and unstable solutions, quantify their stability and accurately detect bifurcation points (we have included the predicted unstable steady state in Figure 2g).

5 Concluding Remarks

The results presented here constitute our initial attempt to apply neural network analysis to chemical reaction data. These initial examples involved relatively simple

dynamic phenomena. The methods however, should be applicable to more complex dynamic behavior, and the extension to incorporate the dependence on more than one parameters is straight forward. It appears that this approach is indeed capable of producing accurate *nonlinear* dynamic models of processes exhibiting low-order dynamics (this does not necessarily mean simple dynamics!). Predictive models can be used to simulate and/or control a process. We should mention that the training in our examples was performed off-line, i.e. the time series were first obtained, and the training was performed subsequently. On-line training however is also possible, and the development of special purpose hardware for such neural network configurations can find use in real-time chemical process identification and control (see for example [17].

6 Acknowledgements

This work was partially supported by the National Science Foundation (JLH and IGK), the Center for Innovative Technology, Commonwealth of Virginia (JLH), a Packard Foundation Fellowship (IGK) and the Air Force Office of Scientific Research (ASL & RMF). The hospitality of the Center of Nonlinear Studies at Los Alamos is gratefully acknowledged.

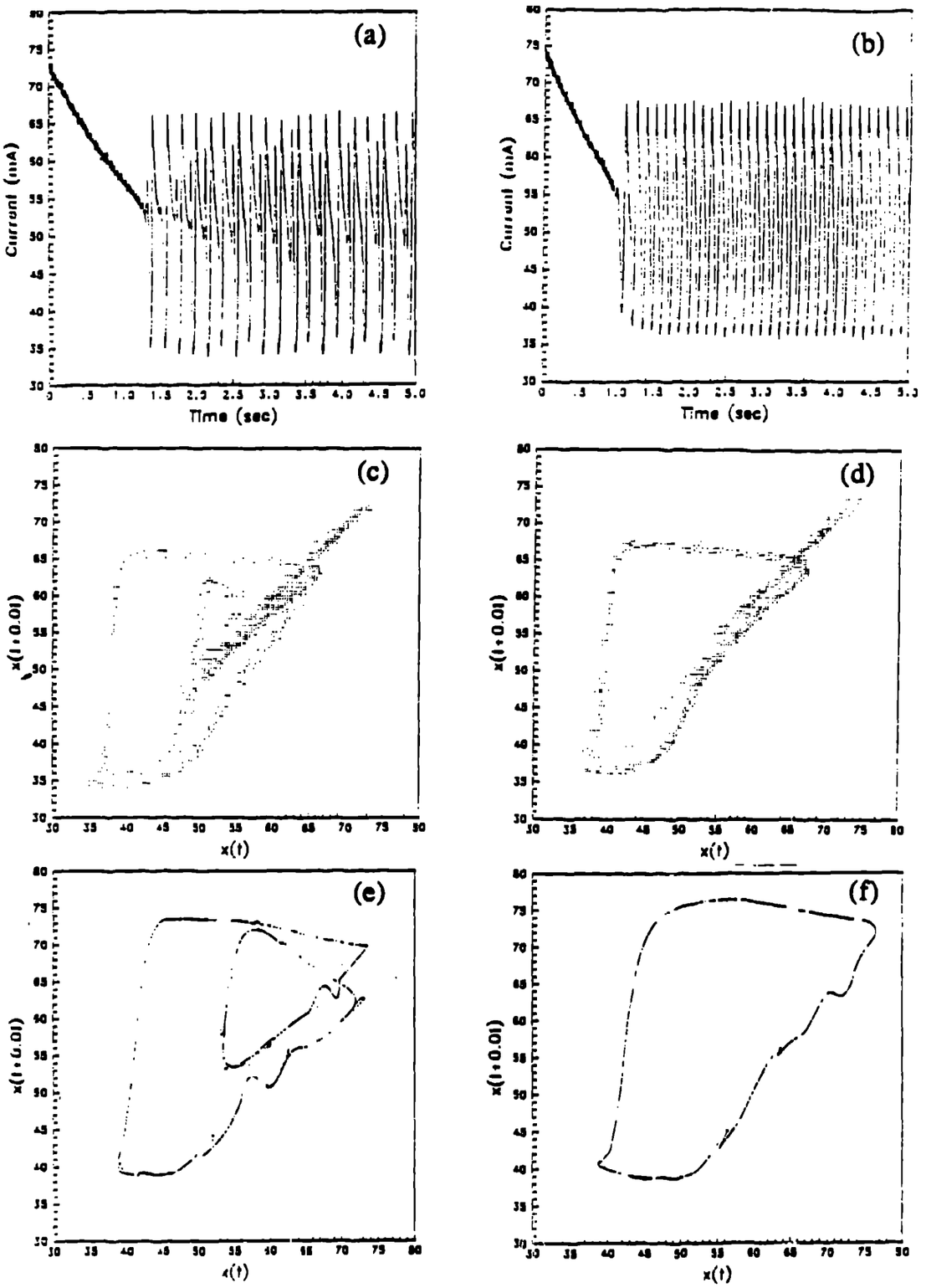

Figure 1: (a-b): Experimental measurements of electrodissolution current vs. time at $769mV$ and $788mV$ respectively. (c-d): The same transients plotted in delay space, leading to a double-loop ($769mV, 2c$) and a single-loop ($788mV, 2d$) attractor. The long straight segment corresponds to the startup part of the transient. (e-f): Attractors predicted by a two hidden layer network.

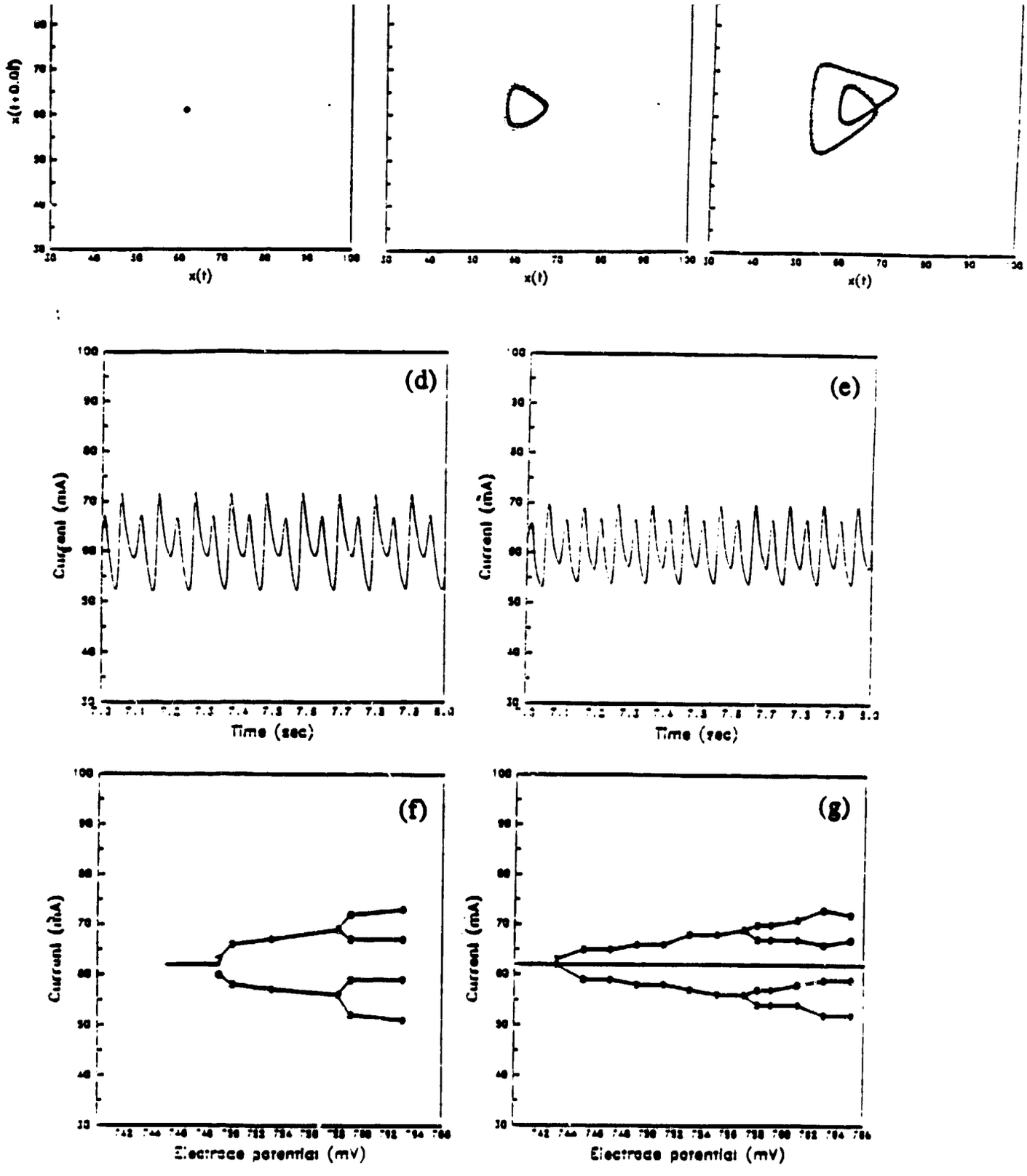

Figure 2: (a,b,c): Phase portraits reconstructed in delay space using experimental measurements at $748mV, 758mV$ and $759mV$ respectively. Segments of time series at $759mV$: experimental (d), and predicted by iterating a pruned radial basis function network (e). Bifurcation diagrams: experimental (f), and predicted using the same radial basis function network (g).

References

[1] M.R. Bassett and J.L. Hudson. Experimental evidence for period-doubling of tori during an electrochemical reaction. *Physica D*, 35, 1989.

[2] M.R. Bassett and J.L. Hudson. Quasiperiodicity and chaos during an electrochemical reaction. *J. Phys. Chem.*, 93, 1989.

[3] M.R. Bassett and J.L. Hudson. S'ilnikov chaos during copper electrodissolution. *J. Phys. Chem.*, 92, 1988.

[4] B.J. Bayly, I. Goldhirsch, and S.A. Orszag. *Independent Degrees of Freedom of Dynamical Systems.* Technical Report, Applied and Computational Mathematics, Princeton U. preprint, 1987.

[5] M. Casdagli. Nonlinear prediction of chaotic time series. *Physica D*, 35, 1989.

[6] J.P. Crutchfield and B.S. McNamara. Equations of motion from a data series. *Complex Systems*, 1:417–452, 1987.

[7] J. D. Farmer and J. J. Sidorowich. Predicting chaotic time series. *Physical Review Letters*, 59(8):845–848, 1987.

[8] J.D. Farmer and J.J. Sidorowich. Exploiting chaos to predict the future and reduce noise. In Y.C. Lee, editor, *Evolution, Learning and Cognition*, World Scientific, 1988.

[9] A.M. Fraser and H.L. Swinney. Independent coordinates for strange attractors from mutual information. *Physical Review*, 33A:1134–1140, 1986.

[10] D. Gabor, W.P. Wilby, and R. Woodcock. A universal nonlinear filter, predictor and simulator which optimizes itself by a learning process. *Proceedings of the IEEE*, 108B:422, 1960.

[11] A. S. Lapedes and R. Farber. Nonlinear signal processing using neural networks: prediction and system modeling. *Techical Report LA-UR 87 2662*, 1987.

[12] A. S. Lapedes and R. M. Farber. How neural nets work. In D.Z. Anderson, editor, *Neural Information Processing Systems*, AIP Press, 1988.

[13] N.H. Packard, J.P. Crutchfield, J.D. Farmer, and R.S. Shaw. Geometry from a time series. *Physical Review Letters*, 45:712–716, 1980.

[14] M.J.D. Powell. *Radial Basis Function for Multivariable Interpolation: A Review.* Technical Report, University of Cambridge, 1985.

[15] F. Takens. Detecting strange attractors in fluid turbulence. In D. Rand and L.S. Young, editors, *Dynamical Systems and Turbulence*, Springer-Verlag, Berlin, 1981.

[16] B. Widrow and S. Stearns. *Adaptive Signal Processing.* Prentice Hall Inc., 1985.

[17] B.D. Ydstie. Forecasting and control using adaptive connectionist networks. 1989. Comp. Chem. Eng.